Proximodistal Exploration in Motor Learning as an Emergent Property of Optimization

Freek Stulp, Pierre-Yves Oudeyer

Proximodistal Exploration in Motor Learning as an Emergent Property of Optimization

Freek Stulp[1,2,3] and Pierre-Yves Oudeyer *[1,2]

[1]Inria, France
[2]ENSTA ParisTech, Université Paris-Saclay, France
[3]German Aerospace Center (DLR), Germany

Abstract

To harness the complexity of their high-dimensional bodies during sensorimotor development, infants are guided by patterns of freezing and freeing of degrees of freedom. For instance, when learning to reach, infants free the degrees of freedom in their arm proximodistally, i.e. from joints that are closer to the body to those that are more distant. Here, we formulate and study computationally the hypothesis that such patterns can emerge spontaneously as the result of a family of stochastic optimization processes (evolution strategies with covariance-matrix adaptation), without an innate encoding of a maturational schedule. In particular, we present simulated experiments with an arm where a computational learner progressively acquires reaching skills through adaptive exploration, and we show that a proximodistal organization appears spontaneously, which we denote PDFF (ProximoDistal Freezing and Freeing of degrees of freedom). We also compare this emergent organization between different arm morphologies – from human-like to quite unnatural ones – to study the effect of different kinematic structures on the emergence of PDFF.
Keywords: human motor learning; proximo-distal exploration; stochastic optimization; modelling; evolution strategies; cross-entropy methods; policy search; morphology.

Research highlights.

- We propose a general, domain-independent hypothesis for the developmental organization of freezing and freeing of degrees of freedom observed both in infant development and adult skill acquisition, such as proximo-distal exploration in learning to reach.
- We introduces a computational model based on basic principles of stochastic optimization, and show how proximodistal freezing and freeing of degrees of freedom arises as an emergent property of this model.
- We analyze the influence of human arm structure on the patterns of freezing and freeing of degrees of freedom in simulated reaching tasks.

1 Introduction

As Bernstein emphasized (Bernstein, 1967), a great mystery in infant motor development is to understand how they can learn motor skills efficiently given a a complex non-linear body with many degrees of freedoms. Robots face the same problems, and this issue has similarly been the object of many studies in the recent years (Vijayakumar, D'souza, & Schaal, 2005; Baranes & Oudeyer, 2011; Kober & Peters, 2011; Baranes & Oudeyer, 2013; Stulp & Sigaud, 2013). Learning motor skills involves experimenting with one's own body under limited time resources, and thus only a small fraction of physically possible movements can be sampled

*Corresponding author: Pierre-Yves Oudeyer (Postal address: Inria, 200, avenue de la vieille tour, 33405 Talence, France; email: pierre-yves.oudeyer@inria.fr)

within the first years of life. Thus, as argued for example in (Berthier, Clifton, McCall, & Robin, 1999) and theoretically analyzed from a machine learning perspective (Oudeyer, Baranes, & Kaplan, 2013), learning strategies based on simple forms of trial and error cannot lead to efficient learning in such contexts.

Several strands of research have studied families of mechanisms that could constrain and guide motor learning processes. In particular, Bernstein established a motor development perspective based on staged learning processes where some degrees of freedoms were first frozen, transforming a complex learning problem in a simpler one, and then progressively freed, allowing the learner to take advantage of the full potential of its body (Bernstein, 1967). A number of experimental studies allowed to confirm this perspective. For example, Berthier et al. (Berthier et al., 1999) showed that the development of early reaching in infants (Bertenthal & von Hofsten, 1998) followed a proximodistal structure, where infants first learnt to reach by freezing the elbow and the hand, while varying shoulder and trunk movements, and then progressively used more distal joints of the elbow and hand. Studies in adult motor skill acquisition showed similar patterning of freezing and freeing of degrees of freedom, applied to the acquisition of racket skills (Southard & Higgins, 1987), soccer (Hodges, Hayes, Horn, & Williams, 2005) or skiing (Vereijken, Emmerik, Whiting, & Newell, 1992). Other experimental observations have shown the complexity and context-dependance of this form of patterning, where for example infant reaching with different postural constraints could lead to higher use of elbow with respect to the shoulder (Thelen et al., 1993).

Several hypotheses explaining the underlying mechanisms leading to such staged motor learning schedules were formulated so far. For example, Berthier et al.(Berthier et al., 1999) suggested that these learning schedules could be innate and due to the progressive neuromuscular development, where physiological maturation of motor neurons along the corticospinal tract (Kuypers, 1981; Jansen & Fladby, 1990) could potentially lead to an initial limitation in the control of distal degrees of freedom. Yet, the extent to which physiological maturation can constrain motor exploration is still unclear in the infant (Adolph & Berger, 2005), and does not provide an explanation of the underlying mechanisms which drive freezing and freeing of degrees of freedom in adult motor learning.

In this article, we formulate, explore and analyze another (possibly complementary) hypothesis from a computational modelling perspective. This hypothesis is formulated within the optimal control framework of motor learning, where the learner uses exploration to find a motor program which minimizes a given cost (or maximizes an objective function) (Todorov, 2004; Berthier, Rosenstein, & Barto, 2005). The hypothesis we study states that staged learning schedules with freezing and progressive freeing of degrees of freedom can self-organize spontaneously as a result of the interaction between certain families of stochastic optimization methods (which drive exploration of the learner) with physical properties of the body, and without involving physiological maturation. In particular, we present simulated experiments with a 6-DOF arm where a computational learner progressively acquires reaching skills (i.e. minimizing a cost to reach), and we show that a proximodistal organization appears spontaneously, which we denote PDFF (Proximo Distal Freezing and Freeing of degrees of freedom). We also compare the emergent structuration as different arm structures are used – from human-like to quite unnatural ones – to study the effect of different kinematic structures on the emergence of PDFF.

In these experiments[1], the reaching task is learned by applying stochastic optimization to optimize the parameters of a movement policy. The algorithm we use – PI^{BB}, a special case of PI^2– is based on covariance matrix adaptation through weighted averaging, which is a concept present in a wide range of optimization frameworks (Arnold, Auger, Hansen, & Ollivier, 2011; Rubinstein & Kroese, 2004; Stulp & Sigaud, 2013). Covariance matrix adaptation allows the algorithm to determine dynamically the appropriate exploration magnitude and direction for each joint in order to progress fastest towards the goal at any given point in the development. In the context of PDFF, increasing and decreasing the exploration corresponds to *freeing* and *freezing* joints respectively.

However, to our knowledge, methods for exploration through covariance matrix adap-

[1]The Matlab code used to generate and visualize the results in this article is available as open source, and can be downloaded here: `https://github.com/stulp/dmp_bbo/archive/proximodistalmaturation.zip`

tation were so far analyzed only from an engineering perspective and in terms of speed to find optimal controllers. Here, on the contrary, we use these general methods as tools to modeling processes of exploration during motor learning in infants and study the patterns of freeing and freezing of DOFs that they generate. Preliminary work in this direction was presented in (Stulp & Oudeyer, 2012b, 2012a), but was based on more complex and specific optimization algorithms, did not include detailed analysis of results, and did not study how different morphologies of the body impacted the resulting patterns of exploration.

Here, we use a simple and generic form of covariance matrix adaptation – PI^{BB}– and study how it spontaneously generates PDFF exploration patterns in the context of several arm morphologies. In the two analysis sections, we further study these results by considering the effect of joints on the cost in a static context. We first perform a sensitivity analysis by quantifying the effect of perturbing individual joint on the main cost component. We then analyze the interactions between joints by determining the effects of perturbing distal joints in the context of perturbations to proximal joints. The results of this analysis provide a deeper understanding of *why* PDFF arises during the stochastic optimization.

2 Limitations of Prior Research

Many computational models have studied how prewired stages or patterns of freezing and freeing of degrees of freedom could contribute or hinder learning of motor skills in high-dimensions. Some studies considered the impact of alternation of freezing and freeing phases upon robot learning of swinging skills (Berthouze & Lungarella, 2004), studied how the pace of the sequencing of discrete stages (Bongard, 2010; Grupen, 2003; Lee, Meng, & Chao, 2007) or of the continuous increase of explored values of DOFs along a proximodistal scheme (Baranes & Oudeyer, 2011) could be adaptively and non-linearly controlled by learning progress and lead to efficient motor learning in high-dimensional robots. Other related models have explored how the progressive freeing of degrees of freedom in the perceptual space (Nagai, Asada, & Hosoda, 2006; French, Mermillod, Quinn, Chauvin, & Mareschal, 2002), in the environment (Uchibe, Asada, & Hosoda, 1998), or in the structure of neural networks for learning abstractions (Elman, 1993; Westermann et al., 2007) could guide the acquisition of sensorimotor and cognitive skills.

In all these models, the global scheduling of freezing and freeing degrees of freedom is encoded by the engineer (but the rhythm of progression from stage to stage can be adaptive as in (Baranes & Oudeyer, 2011; Lee et al., 2007)). Some models have explored explicitly the evolutionary mechanisms that could generate and select such innate maturational schedules (Cangelosi, 1999; Matos, Suzuki, & Arita, 2007).

A related model is presented in Schlesinger et al. (Schlesinger, Parisi, & Langer, 2000). It is most similar to ours in that it also uses a kinematically simulated arm, and explores how evolutionary-like stochastic optimization methods can lead "several constraints to appear to fall out as a consequence of a relatively simple trial-and-error learning algorithm" (Schlesinger et al., 2000), one of them being the locking of joints. Movement policies are represented as four-layer feedforward neural networks, which are trained through evolutionary learning. In terms of the experimental setup, one main difference to our work is that we consider higher-dimensional systems – 10 DOF instead of 3DOF – and use one learning agent instead of a population of 100. A second difference is that we employ a different family of stochastic optimization techniques, which has more flexibility in that it can dynamically update ranges of exploration during single agent learning. While the model in Schlesinger et al. (Schlesinger et al., 2000) only accounted for the freezing of some degrees of freedom as a result of optimization, the flexibility of our learning model allows us to find the entire developmental pattern outlined by Bernstein (Bernstein, 1967): freezing of degrees of freedom followed by progressive and ordered freeing of degrees of freedom. We also consider various arm morphologies to show how the emergent scheduling adapts to the peculiarities of a given kinematic structure.

3 Learning to Reach with Stochastic Optimization

The methods, results and discussions of the experiments are distributed over three sections, corresponding to the three experiments conducted. In this first section, we describe an experiment in which a parameterized policy generates reaching movements, where the parameters of the policy are optimized through stochastic optimization.

3.1 Methods

3.1.1 Arm Model

The evaluation task in this paper consists of a kinematically simulated arm with $M = 6$ degrees of freedom, and a normalized length of 1. To study the effect of different kinematic structures on maturation, we use three sets of relative link lengths, depicted at the bottom of Figure 1: 1) typical relative link lengths of a human arm; 2) equidistant link lengths; 3) 'inverted' human arm, i.e. with short link lengths first.

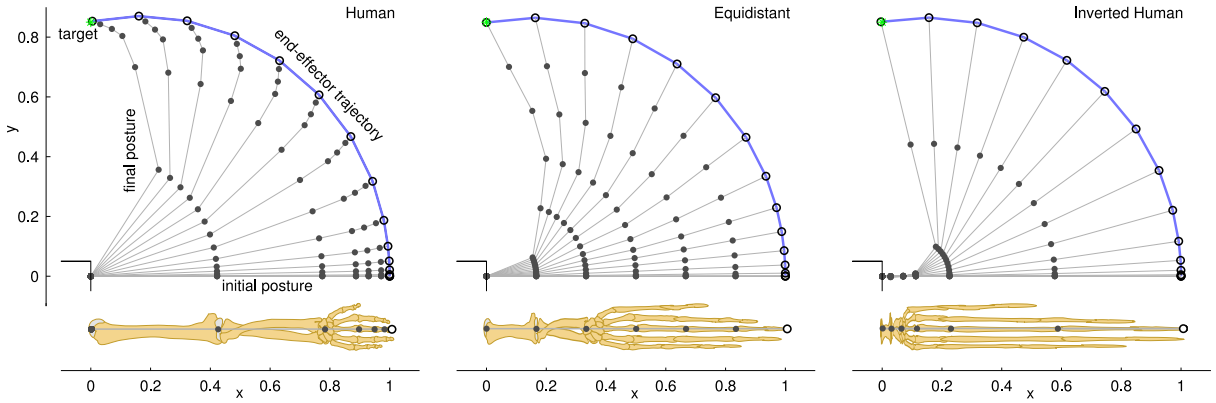

Figure 1: Visualization of the task, which is to reach for a specific target location, in this case $[0, 0.85]$. The arm starts horizontally, and the movement to the target is visualized with a stroboscopic snapshots.

3.1.2 Task Specification

The main aim of the task is to reach for a specific target with a 0.5s movement, visualized in Figure 1. The angles and angular velocities of all joints are initially 0, which corresponds to a completely stretched 'horizontal' arm. The cost function in equations (1)-(2) consists of three parts, expressing different criteria to be optimized during the learning process:

Terminal cost $||\mathbf{x}_{t_N} - \mathbf{x}^g||^2$. The distance between the 2-D Cartesian coordinates of the end-effector ($\mathbf{x}_{t_N}$) at the end of the movement at t_N, and the goal $\mathbf{x}^g$. This expresses that we want to reach to the target $\mathbf{x}^g$ as closely as possible.

Terminal cost $\max(\mathbf{q}_{t_N})$. A cost that corresponds to the largest angle over all the joints at the end of the movement. This expresses an end-state comfort effect (Cohen & Rosenbaum, 2004).

Immediate cost r_t The immediate costs at each time step r_t in (2) penalize joint accelerations. The weighting term $(M + 1 - m)$ penalizes DOFs closer to the origin, the underlying motivation being that wrist movements are less costly than shoulder movements for humans, cf. (Theodorou, Buchli, & Schaal, 2010)[2].

$$\phi_{t_N} = 10^2 ||\mathbf{x}_{t_N} - \mathbf{x}^g||^2 + \max(\mathbf{q}_{t_N}) \qquad \text{Terminal cost} \qquad (1)$$

$$r_t = 10^{-5} \frac{\sum_{m=1}^{M} (M + 1 - m)(\ddot{q}_{t,m})^2}{\sum_{m=1}^{M} (M + 1 - m)} \qquad \text{Immediate cost} \qquad (2)$$

[2]This cost term was taken from (Theodorou et al., 2010). In the context of this paper, it cannot be the reason for the PDFF we shall see in later sections. Rather than favoring PDFF, this cost term actually works *against* it, as proximal joints are penalized *more* for the accelerations that arise due to exploration.

The factors 10^2 and 10^{-5} have two purposes: 1) a scaling factor to compensate for different range of values the different cost components have. 2) a weighting factor enabling the prioritization of tasks. The order of priorities is: reach close to the target, achieve end-state comfort, minimize accelerations.

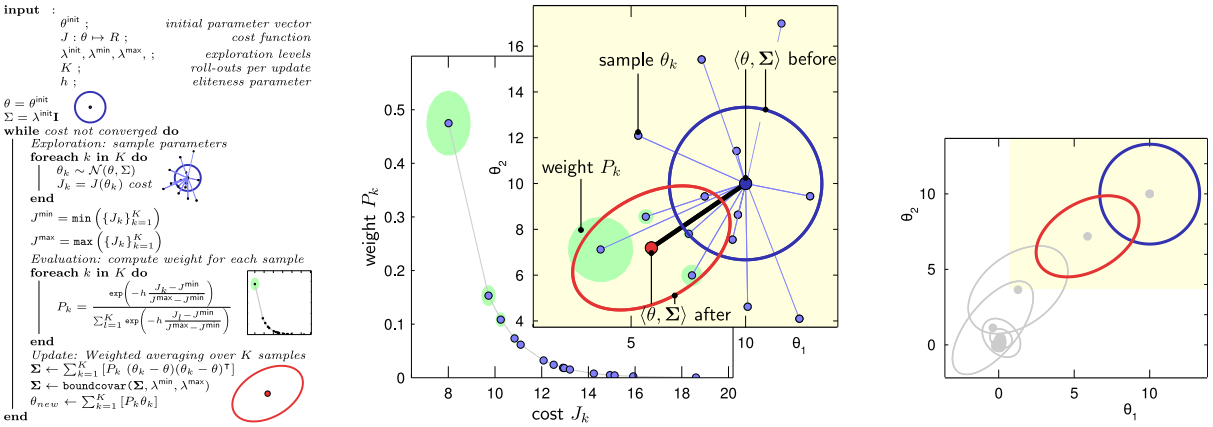

Figure 2: Explanation and visualization of the PI$^{\text{BB}}$ algorithm, using a 2D search space. For illustratory purposes, the cost of a sample θ is simply the distance to the origin: $J(\theta) = ||\theta||$. Left: PI$^{\text{BB}}$ pseudo-code. Center: Visualization of *one* parameter update with PI$^{\text{BB}}$. Right: Evolution of the parameters over *several* updates, demonstrating how the distribution converges towards the minimum $\theta^* = [0, 0]$. The algorithm is initialized by setting the mean and covariance parameters $\langle\theta, \mathbf{\Sigma}\rangle$ to θ^{init} and $\lambda^{\text{init}}\mathbf{I}$ respectively, visualized as a dark blue dot and circle. After this initialization, PI$^{\text{BB}}$ then iteratively updates these parameters with the following steps: 1) *Explore.* Sample K parameter vectors θ_k from $\mathcal{N}(\theta, \mathbf{\Sigma})$, and determine the cost J_k of each sample. In the visualization of our illustratory example task $K = 15$, and the cost $J(\theta)$ is the distance to the origin $||\theta||$, which lies approximately between 8 and 19 in this example. 2) *Evaluate.* Determine the weight (probability) P_k of each sample, given its cost. Essentially, low-cost samples have higher weights, and vice versa. The normalized exponentiation function that maps costs to weights is taken directly from the PI^2 algorithm, and is visualized in the center graph. Larger green circles correspond to higher weights. 3) *Update.* Updating the parameters $\langle\theta, \mathbf{\Sigma}\rangle$ with weighted averaging. In the visualization, the updated parameters are depicted in red. Because low-cost samples (e.g. a cost of 8-10) have higher weights, they contribute more to the update, and θ therefore moves in the direction of the optimum $\theta^* = [0, 0]$.

3.1.3 Policy Representation

The policy representation encodes how movements are generated, by specifying the acceleration profiles of each joint. It is represented as a linear combination of normalized Gaussian basis functions. The acceleration $\ddot{q}_{m,t}$ of the m^{th} joint at time t is determined as in (3), where the parameter vector θ_m represents the weights of joint m.

Intuitively, different basis functions are active at different times during the movement. The first basis function $\Psi_{b=0}(t)$ is most active at the beginning of the movement, and the last $\Psi_{b=B}(t)$ at the end of the movement, with a cascade of basis functions in between. Setting different weights in the parameter vector θ thus leads to different acceleration profiles during the movement. If $\theta = \mathbf{0}$, then there is no acceleration, and thus no movement.

$$\ddot{q}_{m,t} = \mathbf{g}_t^\intercal \theta_m \qquad \text{Acceleration of joint } m \qquad (3)$$

$$[\mathbf{g}_t]_b = \frac{\Psi_b(t)}{\sum_{b=1}^B \Psi_b(t)} \qquad \text{Time-dependent basis functions} \qquad (4)$$

$$\Psi_b(t) = \exp\left(-(t - c_b)^2 / w^2\right) \qquad \text{Kernel} \qquad (5)$$

The centers $c_{b=1...B}$ of the kernels Ψ are spaced equidistantly in the 0.5s duration of the movement, and all have a width of $w = 0.05s$. The number of kernels per joint is $B = 5$. Since we do not simulate arm dynamics, the joint velocities and angles are acquired by integrating the accelerations. The end-effector "hand" position $\mathbf{x}$ is computed with the forward kinematics of the arm.

3.1.4 Policy Improvement through Stochastic Optimization

Stochastic optimization is based on iteratively exploring and updating parameters in a search space θ (θ is the vector of parameters of a movement policy). At each iteration, stochastic optimization algorithms generate K perturbations of the parameter vector $\{\theta_k = \theta + \epsilon_k\}_{k=1}^{K}$, compute the cost J_k for each perturbation, and update the parameters $\theta \rightarrow \theta^{\mathsf{new}}$ based on these costs. This process continues until the costs have converged, or some termination condition is reached. In this article, the parameter space θ corresponds to the parameters of the policy. Optimizing policy parameters is known as *direct policy search*.

The specific stochastic optimization algorithm we use is PI^{BB}, short for "Policy Improvement with Black-Box optimization" (Stulp & Sigaud, 2012). The PI^{BB} algorithm is explained and visualized in Figure 2. We recommend readers to consider Figure 2 in detail, as it is important to understanding the rest of this paper. The main equations from Figure 2 are repeated in (6)-(9).

$$\{\theta_k = \theta + \epsilon_k\}_{k=1}^{K} \qquad \text{Explore by sampling } K \text{ exploratory parameter vectors } \mathcal{N}(\theta, \mathbf{\Sigma}) \tag{6}$$

$$J_k = J(\theta), P_k = f(J_k) \qquad \text{Evaluate by computing } K \text{ costs and map costs to weights} \tag{7}$$

$$\mathbf{\Sigma}^{\mathsf{new}} = \sum_{k=1}^{K} [P_k(\theta_k - \theta)(\theta_k - \theta)^{\intercal}] \qquad \text{Update } \mathbf{\Sigma} \text{ for future exploration with weighted averaging.} \tag{8}$$

$$\theta^{\mathsf{new}} = \sum_{k=1}^{K} [P_k \theta_k] \qquad \text{Update the mean for future sampling with weighted averaging.} \tag{9}$$

The core underlying principle in PI^{BB} relevant to our experiments is using weighted averaging to update the mean θ (9) and covariance matrix $\mathbf{\Sigma}$ (8) of the sampling distribution. It shares this principle with many other *evolution strategies* algorithms such as "Cross-Entropy Methods" (Rubinstein & Kroese, 2004), "Covariance Matrix Adaptation – Evolutionary Strategies" (Hansen & Ostermeier, 2001) (CMA-ES) and "Policy Improvement with Path Integrals" (Theodorou et al., 2010) (PI^2). Hoffman et al. have shown that weighted averaging better explains human motion learning than gradient-based algorithms (Hoffmann, Theodorou, & Schaal, 2008).

The intuitive meaning is that the perturbations that are sampled at each iteration of learning are themselves adaptive, fostering exploration in directions where the cost decreases fastest. Indeed, PI^{BB} is a special case of both PI^2 (without temporal averaging and with covariance matrix adaptation) and CMA-ES (without evolution paths), which are state-of-the-art in direct policy search and black-box optimization respectively.

Here we select PI^{BB} over PI^2 because it is the simplest such algorithm that implements the principle of weighted averaging to update policy parameters; we are interested in studying the formation of staged patterns of freezing and freeing of degrees of freedom (PDFF), not in achieving for instance the fastest possible convergence. For a full explanation of PI^{BB}, and its relationship to PI^2 and CMA-ES, we refer to (Stulp & Sigaud, 2012).

Adaptive Exploration through Covariance Matrix Adaptation Covariance matrix adaptation allows PI^{BB} to automatically adapt the exploration so as to generate more samples in the direction of the minimum. Because this property is important for PDFF, we highlight and illustrate it with the two simple examples in Figure 3.

In the left graph in Figure 3 (repeated from Figure 2), the current parameters $\theta = [10, 10]$ are far from the optimum $\theta^* = [0, 0]$. The samples that are closer to θ^* (to the lower left in the graph) have larger weights (visualized as green circles) than those that are further away from θ^*. Because the new parameters are weighted with these weights when averaging, the mean θ moves in the direction of these high-weight, low-cost samples when updating, bringing θ closer to θ^*. The same principle applies to the covariance matrix, which becomes elongated – its largest eigenvalue λ increases, and corresponds to the eigenvector pointing more in the direction of θ^* (visualized as an arrow). More formally, it follows the natural gradient of the cost with respect to the parameters (Arnold et al., 2011). The effect is that exploration increases, and in the direction of low-cost regions in the parameter space.

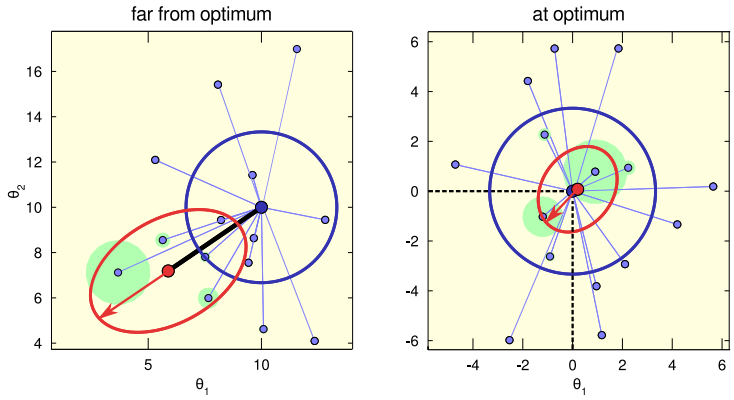

Figure 3: Left: when the current parameters are *far* from the optimum ($\theta^* = [0, 0]$ in this example), the covariance matrix is updated (red ellipse) so that it tends to elongate in the direction of steepest descent, leading to increased exploration along this direction (see arrow). Right: when the current parameters are exactly at the optimum, the covariance matrix tends to shrink, leading to decreased exploration.

In the right graph, the initial covariance matrix is the same, but the current parameters $\theta = [0, 0]$ are now perfectly at the optimum $\theta^* = [0, 0]$. For this reason, samples closer to θ will have a lower cost $J(\theta) = ||\theta||$, and thus a higher weight. Note the larger green circles are all close to the center. Therefore, θ hardly moves after updating; this is desirable, because θ is already at the optimum. However, the covariance matrix shrinks (see smaller eigenvalue as arrow) because closer samples have higher weights. Therefore, exploration decreases in all directions.

This adaptive exploration behavior may also be observed in the right graph in Figure 2; in the first few updates the covariance matrix elongates towards the optimum $\theta^* = [0, 0]$, but once it is reached, the covariance matrix shrinks and exploration decreases. This behavior is especially apparent in this idealized example, which uses only a 2-D search space and a very benign cost function; however, the general principle also applies to high-dimensional spaces and discontinuous cost functions, as demonstrated in (Stulp, 2012).

3.1.5 Application of PI$^{\text{BB}}$ to the Reaching Task

In this article, we apply PI$^{\text{BB}}$ to the parameters of the policy representation previously described. Each joint has its own parameters θ_m and covariance matrix $\mathbf{\Sigma}_m$, which are iteratively updated with PI$^{\text{BB}}$. The input parameters of PI$^{\text{BB}}$ are set as follows. The initial parameter vector is $\theta^{\text{init}} = \mathbf{0}$, which means the arm is completely stretched, and not moving at all over time. The initial and minimum exploration magnitude of each joint m is set to $\lambda^{\text{init}} = \lambda^{\text{min}} = 0.05$. The number of trials per update is $K = 20$, and the eliteness parameter is $h = 10$, the default value suggested by (Theodorou et al., 2010)[3] A separate stochastic

[3]High values of the eliteness h lead only a few samples to contribute to the weighted averaging. $h = 0$ would give all samples equal weight independent of the cost, and no learning would occur. As (Hansen & Ostermeier, 2001) note: "In general, the selection related parameters [such as h] are comparatively uncritical and can be chosen in a wide range without disturbing the adaptation procedure." In fact, the same parameter settings have been used in entirely different domains, for instance to optimize robot control policies (Theodorou et al., 2010).

optimization session was run 10 times for each of the 20 target points, i.e. 200 sessions per arm structure.

The exploration magnitude λ_m of a particular joint m at some point during the learning process is defined as the maximum eigenvalue of the covariance matrix $\mathbf{\Sigma}_m$. Initially, λ_m is λ^{init}, because the all eigenvalues of the initial diagonal $\mathbf{\Sigma} = \lambda^{\text{init}}\mathbf{I}$ are λ^{init}. The length of the two arrows in Figure 3 visualize λ_m for non-diagonal covariance matrices, which may arise as $\mathbf{\Sigma}$ is updated.

In the context of this paper, we consider joints that have low exploration magnitudes λ_m to be 'frozen', whereas those with high λ_m are 'free'.

3.2 Results

Figure 4 presents the results of applying PI^{BB} to the three different arm structures. Each graph plots the total exploration magnitude over all joints $\sum_{m=1}^{M} \lambda_m$ (thick yellow/black line) and the relative exploration magnitude $\lambda_m / \sum_{m=1}^{M} \lambda_m$ per joint (colored patches) as learning progresses over 100 updates, which corresponds to $2000 = 100 \cdot K$ roll-outs. All values are averaged over the 20 target points and 10 optimization sessions per target point. The thick vertical bars indicates when a joint reached its maximum relative exploration magnitude (position on x-axis), as well as the magnitude itself (the height of the bar). For example, for the human arm structure (top graph), the first joint achieves a maximum relative exploration magnitude of 0.56 at update 7. This means that 56% of the exploration is accounted for by only the most proximal (first) joint.

To investigate the robustness of the method against initial conditions, Figure 5 plots the variability around the mean of the exploration in the first joint in the human arm for the 200 learning sessions for this arm type. In generating this figure, we noticed that the exact onset of increasing exploration (freeing degrees of freedom) is influenced by the stochastic nature of the optimization algorithm. To factor this out, we have applied dynamic time warping (Sakoe & Chiba, 1978) to the exploration curves of the 200 individual learning sessions before computing the mean and variance.

3.3 Discussion

For all arm structures, we see that the total exploration (thick black/yellow line) initially increases, indicating that DOFs are globally freed. After achieving a maximum total exploration at around update 20-25, exploration then decreases again once the task has been learned. This behavior is a direct consequence of the adaptive exploration described in Figure 3, and is also observed in (Stulp & Oudeyer, 2012b; Stulp, 2012).

The relative exploration magnitude between the joints, however, shows quite a different development for the different arm structures. For the human arm (top graph), the most proximal joint is already responsible for more than 50% of the exploration (0.56) after 7 updates. This joint has been freed, whereas the others are frozen; the three distal joints account for less than 20% of exploration at update 7. For the human arm, we see that joints 1, 2, 3 achieve their maximum relative exploration of 0.56, 0.42 and 0.26 at updates 7, 18 and 27 respectively. In conclusion, we clearly see that the first three joints are freed in a proximodistal order[4].

For the equidistant arm, the order in which the joints achieve their maxima is 2,3,4,1,5,6. Thus, apart from the most proximal joint, we again see a proximodistal freeing of joints. When considering the maximum relative magnitudes of the exploration (vertical bars), we see however that the freeing/freezing of joints is much less less pronounced. None of the maxima exceeds 0.3, which is in contrast with the human arm, where the most proximal joint is responsible for 0.56 if the exploration. For the equidistant arm, the exploration is thus spread out over the joints much more, rather than being focused in only one or several joints.

Finally, for the inverted human arm, the order in which joints are freed is 3,4,5,6,1,2. This time, the bulk of the exploration is being done by joint 6, which accounts for more

[4]Note that the 6th joint achieves its maximum at update 1. This is not because it is freed very early, but rather because it is almost frozen throughout the learning process, and thus achieves its maximum when the exploration is initially the same for all joints

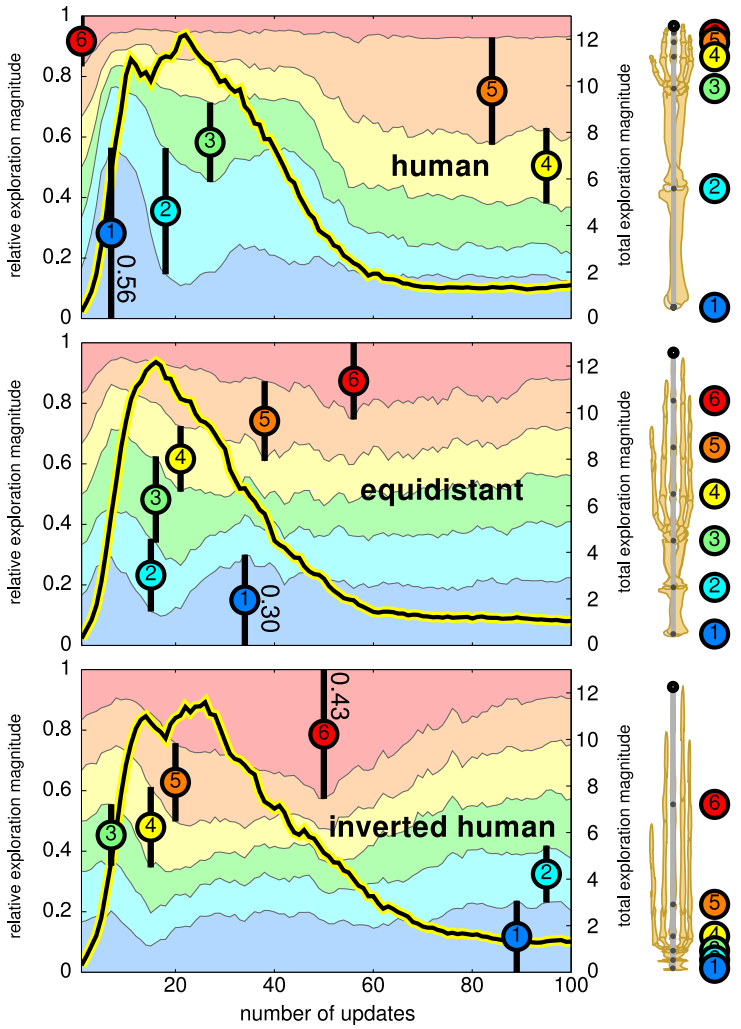

Figure 4: Results of stochastic optimization for the three arm structures.The thick black/yellow line represents the total exploration magnitude (right y-axis) and relative exploration magnitudes of the 6 individual joints as learning progresses (left y-axis).

than 42% of exploration at update 50. The other joints again never exceed 0.3. Thus, for this rather unnatural arm, we see more exploration in the distal joints; only later on do the proximal joints 1 and 2 achieve their maximum values.

In summary, the results show that there is a consistent emergent organization of exploration over time in all arm structures. The PDFF organization is quite pronounced in the human arm, where the order of freeing is 1,2,3,5,4, and the exploration switches most clearly (i.e. high relative magnitudes of exploration) from one joint to the other. It is important to realize that this effect emerges solely from adaptive exploration through covariance matrix adaptation, and the order and/or stages in which degrees-of-freedom are freed is not pre-defined, as in for instance (Berthouze & Lungarella, 2004; Bongard, 2010; Baranes & Oudeyer, 2011).

4 Analysis: Individual Joints

In this section, we analyze how and why PDFF arises when applying stochastic optimization, and how and why this depends on the arm structure. We perform the analysis in a static context – static because we do not perturb the parameters of the policy that determines the joint angles over time, but rather perturb the joint angles directly without a temporal component. This analysis helps to understand why PDFF arises within an optimization context.

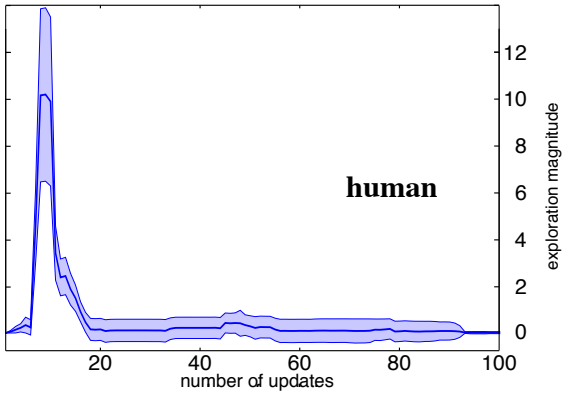

Figure 5: Mean and standard deviation in the exploration magnitude of the first joint in the human arm at each update, after aligning the 200 sessions for this arm with dynamic time warping. This figure illustrates that the variance is higher when exploration is highest, as is to be expected. The variances for the other joints, not plotted here, are lower.

Here, we consider the effect that perturbations of individual joints have on the cost through sensitivity analysis. Sensitivity analysis aims at *"providing an understanding of how the model response variables respond to changes in the input."* (Saltelli, Chan, & Scott, 2000). We use sensitivity analysis to investigate how the variation in individual joint angles – the input – influences the variation in the cost – the response variables. This provides a first indication of why PDFF arises.

4.1 Methods

In the default posture, all joint angles are zero. This posture is perturbed by setting one of the 6 joint angles to $\frac{\pi}{10}$. The 6 possible perturbations, one for each joint, are visualized in the top row of Figure 6. For the default and perturbed configuration, we then compute the distance of the end-effector to the target $||\mathbf{x}_{t_N} - \mathbf{x}^g||$. Because this is a static context there are no joint accelerations, and the immediate costs (2) are not included. The lower row plots the difference in cost between the outstretched arm (where all joints are 0), and the slightly bent arm (where one joint angle is $\frac{\pi}{10}$). To acquire a value that is representative for the whole workspace, the differences in the lower row of Figure 6 is the average over the 20 target positions $\mathbf{x}^g$ depicted in the top row of Figure 6.

4.2 Results

For all arm configurations, we see that proximal joints lead to a higher average difference in the distance to the target than more distal ones. This should not come as a surprise, as rotating more proximal joint leads to smaller movement in the end-effector space, and it is the end-effector space that determines the distance to the target. As a consequence, the

same magnitude of perturbation will lead to a larger difference in cost for more proximal joints.

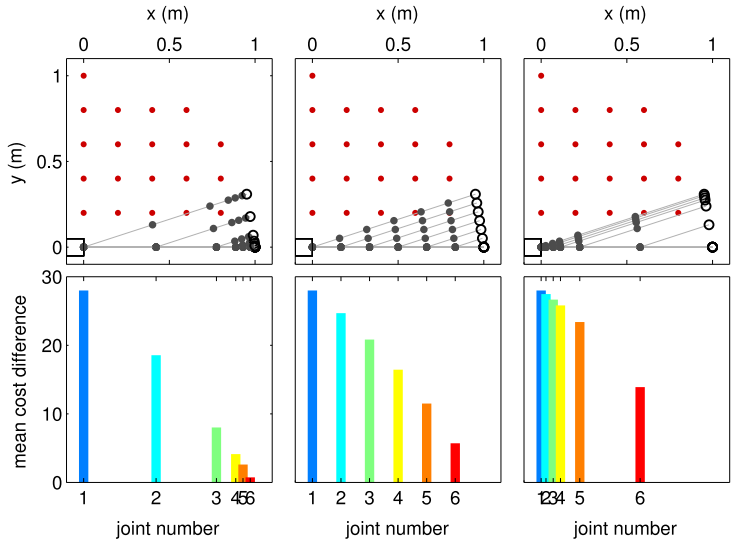

Figure 6: Results of the sensitivity analysis. For all arm configurations, we see that proximal joints lead to a higher average difference in the distance to the target than more distal ones.

4.3 Discussion

The goal of stochastic optimization is to minimize costs through exploring and updating in parameter space. The results in Figure 6 demonstrate that perturbing proximal joints leads to larger differences in costs than distal joints. Therefore, an optimizer can be expected to minimize costs more quickly if it initially focuses exploration on proximal joints, rather than distal ones. This may be an explanation why exploration is larger in more proximal joints, but does not explain why distal joints are not also freed. This is the aim of our second analysis, which now follows.

5 Analysis: Interactions Between Joints

Whereas the previous section on sensitivity analysis considered joints *individually*, we now turn to the *interaction* between pairs of joints. We especially focus on how perturbations in proximal joints affect the influence of perturbations in more distal joints on the cost.

5.1 Methods

Pairs of joints are considered. For the more proximal joint, two perturbations $\{P1, P2\}$ are sampled from $\mathcal{N}(0, \frac{\pi}{10})$. For each perturbation of the proximal joint, the distal joint is perturbed twice, also by sampling from $\mathcal{N}(0, \frac{\pi}{10})$. This leads to the four arm configurations in Figure 7. The question we ask for these four configurations is: does the perturbation of the distal joint change the cost ranking? This is the same question underlying *uncertainty handling* in ranked-based evolutionary direct policy search (Heidrich-Meisner & Igel, 2008). Figure 7 depicts examples where the answer is no (left) and yes (right). The question is asked for 100 different samples, and for each of the 20 target points. The average of these 2000 values represents the ratio that the answer was 'no', i.e. a ratio of 1 implies that, no matter how much the distal joint is perturbed, it does not affect the ranking. A value of 0.5 implies that the perturbation of the distal joint affects the ranking half the time.

5.2 Results

Figure 8 depicts the ratio for pairs of joints for all three arm configurations. For proximal joint 1 and distal joint 3 – the case depicted in Figure 7 – this value is 0.89, labeled (A).

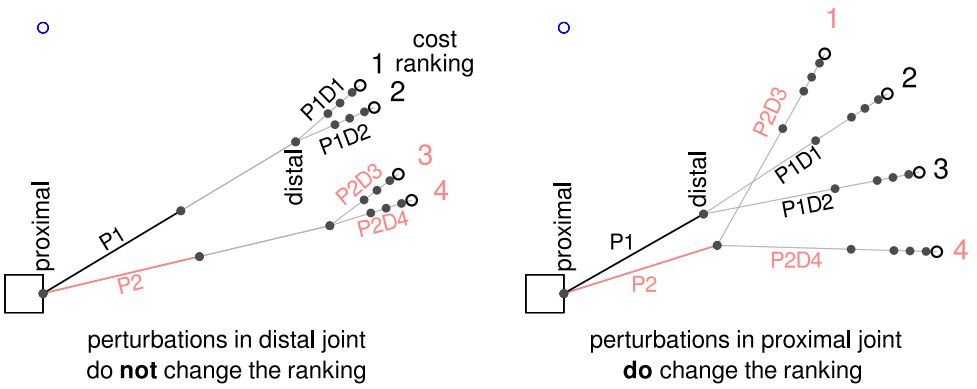

Figure 7: Examples of perturbing pairs of joints for the human arm, and the effect of these perturbations on the cost ranking (numbers at the end of the arm).

Thus, when joint 1 is perturbed, the perturbation of joint 3 affect the cost ranking in only 11% of the samples and target points.

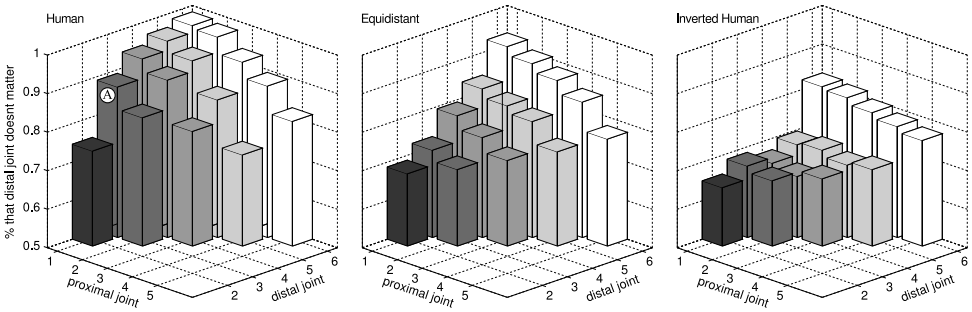

Figure 8: For pairs of joints, the height of the bars represents the ratio that the perturbation of the distal joint does *not* affect the cost ranking of the resulting postures.

5.3 Discussion

The main implication of these results for stochastic optimization is that if we are performing exploration with for instance joint 1, exploration in more distal joints is less relevant to the cost. For the human arm for instance, it is not sensible to explore with joint 6 when exploring with joint 1, because joint 6 only affects the result 2% of the time. Thus, from the point of view of stochastic optimization, joint 6 may well be frozen when searching in joint 1.

It is interesting to see that for the human arm configuration, proximal joints dominate distal joint much more (median=0.89) than an equidistant arm (median=0.79), and even more so for the 'inverted' human arm (median=0.70). Thus, exploring with more distal joints has a lower impact on the cost ranking for the human arm, and thus the effect of PDFF may be expected to be stronger for this configuration. This is confirmed by the empirical results in Figure 4, where PDFF is most pronounced in the human arm configuration.

6 General Discussion and Conclusion

Staged motor learning, and in particular the progressive freezing and freeing of degrees of freedom observed in infant and adult motor exploration, has been argued to facilitate human acquisition of high-dimensional motor skills, and was also shown to be efficient for robot motor learning. Several hypotheses explaining the underlying mechanisms leading to such staged motor learning schedules were formulated so far, but have mostly relied on forms of innate scheduling of patterns of freeing and freezing.

Here, in the framework of approximate optimal control, we have studied the hypothesis that staged learning schedules with freezing and progressive freeing of degrees of freedom

can self-organize spontaneously as a result of the interaction between certain families of stochastic optimization methods with the physical properties of the body, and without involving physiological maturation. In particular, we have presented simulated experiments with a 6-DOF arm where a computational learner progressively acquired reaching skills, and we showed that a proximodistal organization appeared spontaneously. We also compared the emergent structuration as different arm structures are used – from human-like to quite unnatural ones – to study the effect of different kinematic structures on the emergence of PDFF. We analyzed further these results through a sensitivity analysis, providing a deeper understanding of *why* PDFF arises during the stochastic optimization.

Parsimony and biological plausibility Overall, this model does not invalidate the hypothesis that an an innate maturational scheduling for freezing and freezing DOFs can be involved in infant motor learning. However, it shows that relatively simple stochastic optimization processes with adaptive exploration, which Bernstein already suggested were at play in infants, can already account for the formation of patterns of staged motor exploration. Yet, while the form of adaptive stochastic optimization we have considered is simple and general, one can wonder whether such mechanisms could be actually implemented in biologically plausible neural networks. These mechanisms are a form of evolutionary optimization algorithms that are based on two complementary principles: the capacity to make variations/mutations of current good solutions (and to select the most useful ones), and the capacity to identify which directions of variation/mutation are currently most improving the current good solutions, involving a form of memory of past explorations. These two principles, and other more complex forms of Darwinian search processes in the brain, have been shown to be neurally plausible by several lines of research (C. T. Fernando, Szathmary, & Husbands, 2012), building on Edelman theory of neuronal group selection (Edelman, 1987), Changeux's theory of synaptic selection and selective stabilization (Changeux, Courrége, & Danchin, 1973), Calvin's replicating activity patterns (Calvin, 1987). In particular, the recently developped Neuronal Replicator Hypothesis (C. Fernando, Goldstein, & Szathmáry, 2010) has shown how various known neuronal physiological mechanisms could implement such general genetic algorithms, including the mechanisms of adaptive exploration that we have been using in the model presented in this paper. An implementation of these mechanisms was shown to work with a neural network using realistic Izhikevich spiking neurons (C. Fernando, Vasas, Szathmáry, & Husbands, 2011). To summarize, the mechanisms used to allow incremental exploration and learning in the model presented here are not only simple and parsimonious, but they are also implementable in realistic neural networks.

Open questions and research perspectives. Our general hypothesis therefore forms a baseline against which more complex, domain-specific hypotheses should be compared. Also, the spontaneous formation of PDFF patterns through stochastic optimization appears to be compatible with observation of the patterns of motor exploration in adult motor learning, where maturational mechanisms have little probability to be at play. This new hypothesis also points to several open questions and new experimental investigations. In adults, several experiments have shown that a structuration of exploration through freezing and freezing of degrees of freedom happened (Southard & Higgins, 1987; Hodges et al., 2005; Vereijken et al., 1992), however to our knowledge these experiments did not systematically study to what extent the freeing or freezing of particular degrees of freedom was correlated to their current usefulness in progressing towards a goal. Also, they considered tasks such as skiing, soccer or racket skills that were culturally known by subjects and thus could involve other mechanisms such as imitation learning, complicating the analysis of the exploration strategies. In infants, it is also an open question to know whether the adaptive exploration mechanisms shown by adults are already at play, or whether they are not yet in place and exploration is rather controlled by maturational mechanisms such as myelination.

To address these open questions, one could imagine using experimental setups in which human subjects have to learn and explore new sensorimotor mapping that are highly different from what they already know, and where the relation between degrees of freedom and their usefulness for progressing towards the goal can be controlled systematically. Miard et al. (Miard et al., 2014) have proposed an experimental setup that could be used with adults in this context (but was not used for these specific questions so far): subjects have to learn how to control an abstract visual shape on a screen by using movements of their body as measured by a 3D camera (and the mathematical form of the relation between

body movements and the abstract visual shape can be changed systematically). Similar setups could be imagined for infants, taking inspiration from the famous Rovee-Collier task (Rovee-Collier, 1999) (developped for studying other questions): infants' movements (arms, legs) could be tracked with sensors and used to control the intensity/frequency of a sound or color of a light, through a mapping where the relation between degrees of freedom and their impact on the sound/light could be systematically changed to cancel out possible effects of myelination in the structuration of exploration. For example, it could be possible to program an inverse proximodistal relationship between arm movements and the sound/light: movements of the tip of the arm could be made to have more impact on the sound/light than movements of the shoulder. In such a case, observing an infant exploration with a corresponding inverse proximodistal law would reinforce our hypothesis that adaptive exploration plays an important role, while observing a standard proximodistal exploration would invalidate the hypothesis that such a mechanism is present or can have a leading role in exploration.

Complementarity with other mechanisms It is also an open question to undertand how such adaptive exploration based on stochastic optimization could interact with other mechanisms guiding exploration such as maturation through myelination, imitation/social guidance, and intrinsic motivation. Several models in robotics have began to explore these links. Baranes and Oudeyer (Baranes & Oudeyer, 2011) have studied the efficiency of combining stochastic optimization to reach goals with maturational mechanism which progressively grow the limits within which stochastic optimization can physically explore, showing an increase in efficiency from a machine learning point of view. Several works have shown how human demonstration of movements could bootstrap this optimization process (e.g. (Stulp, Herlant, Hoarau, & Raiola, 2014), or how humans can progressively shape subparts of the movements to complement autonomous exploration (Chernova & Thomaz, 2014). Finally, exploration in infants is also highly driven by mechanisms of intrinsic motivation (also called curiosity), where instead of trying to reach a goal imposed by social peers or the experimenter (as in the model presented in this paper), they use intrinsic criteria such as information gain or surprise to set their own goals and choose how to practice these self-selected goals (Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Moulin-Frier, Nguyen, & Oudeyer, 2014). An important side effect of exploring multiple self-selected goals is that transfer learning across goals happens and in turn can shape the selection of future goals and ways to use degrees of freedom to explore them. Several computational architectures of intrinsically motivated learning have used stochastic optimization as the lower-level mechanism to learn how to reach self-selected goals in sensorimotor learning and achieve transfer learning across goals (Baranes & Oudeyer, 2013). These integrated architecture have also shown the self-organization of developmental structure, such as the transition from non-articulated speech sounds to articulated vowels to proto-syllables in models if infant vocal development (Moulin-Frier et al., 2014; Oudeyer & Smith, 2016).

References

Adolph, K. E., & Berger, S. E. (2005). Physical and motor development. *Developmental science: An advanced textbook*, *5*, 223–281.

Arnold, L., Auger, A., Hansen, N., & Ollivier, Y. (2011). *Information-geometric optimization algorithms: A unifying picture via invariance principles* (Tech. Rep.). INRIA Saclay.

Baranes, A., & Oudeyer, P.-Y. (2011). The interaction of maturational constraints and intrinsic motivations in active motor development. In *Ieee international conference on development and learning.*

Baranes, A., & Oudeyer, P.-Y. (2013, January). Active Learning of Inverse Models with Intrinsically Motivated Goal Exploration in Robots. *Robotics and Autonomous Systems*, *61*(1), 69-73. Retrieved from `http://hal.inria.fr/hal-00788440` doi: 10.1016/j.robot.2012.05.008

Bernstein, N. (1967). *The coordination and regulation of movements.* Pergamon.

Bertenthal, B. I., & von Hofsten, C. (1998). Eye, head and trunk control: The foundation for manual development. *Neuroscience and Biobehavioral Review*, *22*, 515-520.

Berthier, N. E., Clifton, R., McCall, D., & Robin, D. (1999). Proximodistal structure of early reaching in human infants. *Exp Brain Res*.

Berthier, N. E., Rosenstein, M. T., & Barto, A. G. (2005). Approximate optimal control as a model for motor learning. *Psychological review*, *112*(2), 329.

Berthouze, L., & Lungarella, M. (2004). Motor skill acquisition under environmental perturbations: On the necessity of alternate freezing and freeing degrees of freedom. *Adaptive Behavior*, *12*(1), 47-63.

Bongard, J. C. (2010, January). Morphological change in machines accelerates the evolution of robust behavior. *Proceedigns of the National Academy of Sciences of the United States of America (PNAS)*.

Calvin, W. H. (1987). The brain as a darwin machine. *Nature*, *330*, 33–34.

Cangelosi, A. (1999). Heterochrony and adaptation in developing neural networks. In W. B. et al. (Ed.), *Proceedings of the genetic and evolutionary computation conference* (pp. 1241–1248). San Francisco, CA: Morgan Kaufmann.

Changeux, J.-P., Courrége, P., & Danchin, A. (1973). A theory of the epigenesis of neuronal networks by selective stabilization of synapses. *Proceedings of the National Academy of Sciences*, *70*(10), 2974–2978.

Chernova, S., & Thomaz, A. L. (2014). Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *8*(3), 1–121.

Cohen, R. G., & Rosenbaum, D. A. (2004). Where grasps are made reveals how grasps are planned: generation and recall of motor plans. *Exp Brain Res*, *157*(4), 486-495. doi: 10.1007/s00221-004-1862-9

Edelman, G. M. (1987). *Neural darwinism: The theory of neuronal group selection.* Basic Books.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71-99.

Fernando, C., Goldstein, R., & Szathmáry, E. (2010). The neuronal replicator hypothesis. *Neural computation*, *22*(11), 2809–2857.

Fernando, C., Vasas, V., Szathmáry, E., & Husbands, P. (2011). Evolvable neuronal paths: a novel basis for information and search in the brain. *PloS one*, *6*(8), e23534.

Fernando, C. T., Szathmary, E., & Husbands, P. (2012). Selectionist and evolutionary approaches to brain function: a critical appraisal. *Frontiers in computational neuroscience*, *6*, 24.

French, R. M., Mermillod, M., Quinn, P. C., Chauvin, A., & Mareschal, D. (2002). The importance of starting blurry: Simulating improved basic-level category learning in infants due to weak visual acuity. In LEA (Ed.), *Proceedings of the 24th annual conference of the cognitive science society* (p. 322-327). New Jersey.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585–593.

Grupen, R. (2003, August). A developmental organization for robot behavior. In *Proceedings of the third international workshop on epigenenetic robotics.* Boston, MA.

Hansen, N., & Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, *9*(2), 159-195.

Heidrich-Meisner, V., & Igel, C. (2008). Uncertainty handling in evolutionary direct policy search..

Hodges, N. J., Hayes, S., Horn, R. R., & Williams, A. M. (2005). Changes in coordination, control and outcome as a result of extended practice on a novel motor skill. *Ergonomics*, *48*(11-14), 1672–1685.

Hoffmann, H., Theodorou, E., & Schaal, S. (2008). Optimization strategies in human reinforcement learning. *Advances in computational motor control VII. Washington, DC: Society for Neuroscience*.

Jansen, J., & Fladby, T. (1990). The perinatal reorganization of the innervation of skeletal muscle in mammals. *Progress in neurobiology*, *34*(1), 39–90.

Kober, J., & Peters, J. (2011). Policy search for motor primitives in robotics. *Machine Learning*, *84*, 171-203.

Kuypers, H. (1981). Anatomy of the descending pathways. *Comprehensive Physiology*.

Lee, M., Meng, Q., & Chao, F. (2007). Developmental learning for autonomous robots. *Robotics and Autonomous Systems*, *55*(9), 750-759.

Matos, A., Suzuki, R., & Arita, T. (2007). Heterochrony and evolvability in neural network development. *Artificial Life Robotics*, *11*, 175-182.

Miard, B., Rouanet, P., Grizou, J., Lopes, M., Gottlieb, J., Baranes, A., & Oudeyer, P.-Y. (2014). A new experimental setup to study the structure of curiosity-driven exploration in humans. In *Proceedings of icdl-epirob 2014.*

Moulin-Frier, C., Nguyen, S. M., & Oudeyer, P.-Y. (2014). Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Frontiers in Psychology*, *4*, 1006. Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2013.01006 doi: 10.3389/fpsyg.2013.01006

Nagai, Y., Asada, M., & Hosoda, K. (2006, September). Learning for joint attention helped by functional development. *Advanced Robotics*, *20*(10), 1165-1181.

Oudeyer, P.-Y., Baranes, A., & Kaplan, F. (2013, February). Intrinsically Motivated Learning of Real World Sensorimotor Skills with Developmental Constraints. In G. Baldassarre & M. Mirolli (Eds.), *Intrinsically Motivated Learning in Natural and Artificial Systems.* Springer. Retrieved from http://hal.inria.fr/hal-00788611

Oudeyer, P.-Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, *8*(2), 492–502. Retrieved from http://dx.doi.org/10.1111/tops.12196 doi: 10.1111/tops.12196

Rovee-Collier, C. (1999). The development of infant memory. *Current Directions in Psychological Science*, *8*(3), 80–85.

Rubinstein, R., & Kroese, D. (2004). *The cross-entropy method: A unified approach to combinatorial optimization, monte-carlo simulation, and machine learning.* Springer-Verlag.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, *26*, 43–49.

Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity analysis.* Chichester: Wiley.

Schlesinger, M., Parisi, D., & Langer, J. (2000). Learning to reach by constraining the movement search space. *Developmental Science*, *3*, 67-80.

Southard, D., & Higgins, T. (1987). Changing movement patterns: Effects of demonstration and practice. *Research quarterly for exercise and sport*, *58*(1), 77–80.

Stulp, F. (2012). Adaptive exploration for continual reinforcement learning. In *International conference on intelligent robots and systems (iros)* (p. 1631-1636).

Stulp, F., Herlant, L., Hoarau, A., & Raiola, G. (2014). Simultaneous on-line discovery and improvement of robotic skill options. In *International conference on intelligent robots and systems (iros)*.

Stulp, F., & Oudeyer, P.-Y. (2012a, September). Adaptive exploration through covariance matrix adaptation enables developmental motor learning. *Paladyn. Journal of Behavioral Robotics*, *3*(3), 128–135.

Stulp, F., & Oudeyer, P.-Y. (2012b). Emergent proximo-distal maturation through adaptive exploration. In *International conference on development and learning (icdl)*. (**Paper of Excellence Award**)

Stulp, F., & Sigaud, O. (2012). *Policy improvement methods: Between black-box optimization and episodic reinforcement learning.* Retrieved from `http://hal.archives-ouvertes.fr/hal-00738463` (hal-00738463)

Stulp, F., & Sigaud, O. (2013, September). Robot skill learning: From reinforcement learning to evolution strategies. *Paladyn. Journal of Behavioral Robotics*, *4*(1), 49–61.

Thelen, E., Corbetta, D., Kamm, K., Spencer, J. P., Schneider, K., & Zernicke, R. F. (1993). The transition to reaching: Mapping intention and intrinsic dynamics. *Child development*, *64*(4), 1058–1098.

Theodorou, E., Buchli, J., & Schaal, S. (2010). A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, *11*, 3137-3181.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7*(9), 907-915.

Uchibe, E., Asada, M., & Hosoda, K. (1998). Environmental complexity control for vision-based learning mobile robot. In *roceedings of ieee international conference on robotics and automation* (pp. 1865–1870). IEEE Press.

Vereijken, B., Emmerik, R. E. v., Whiting, H., & Newell, K. M. (1992). Free (z) ing degrees of freedom in skill acquisition. *Journal of motor behavior*, *24*(1), 133–142.

Vijayakumar, S., D'souza, A., & Schaal, S. (2005). Incremental online learning in high dimensions. *Neural computation*, *17*(12), 2602–2634.

Westermann, G., Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M. W., & Thomas, M. S. (2007). Neuroconstructivism. *Developmental science*, *10*(1), 75–83.